SOLERS
An Employee Owned Company

Next Generation Cloud Based Ingest & Processing Framework (I&PF) for Environmental Data

2017 AMS Annual Meeting

Rich Baker
Chief Architect
Solers, Inc.
Email: richard.baker@solers.com
Phone: 240-790-3338

Josh Leaverton, Peter MacHarrie, Dan Beall
Solers, Inc.

Dr. Shay Strong
OmniEarth, Inc.

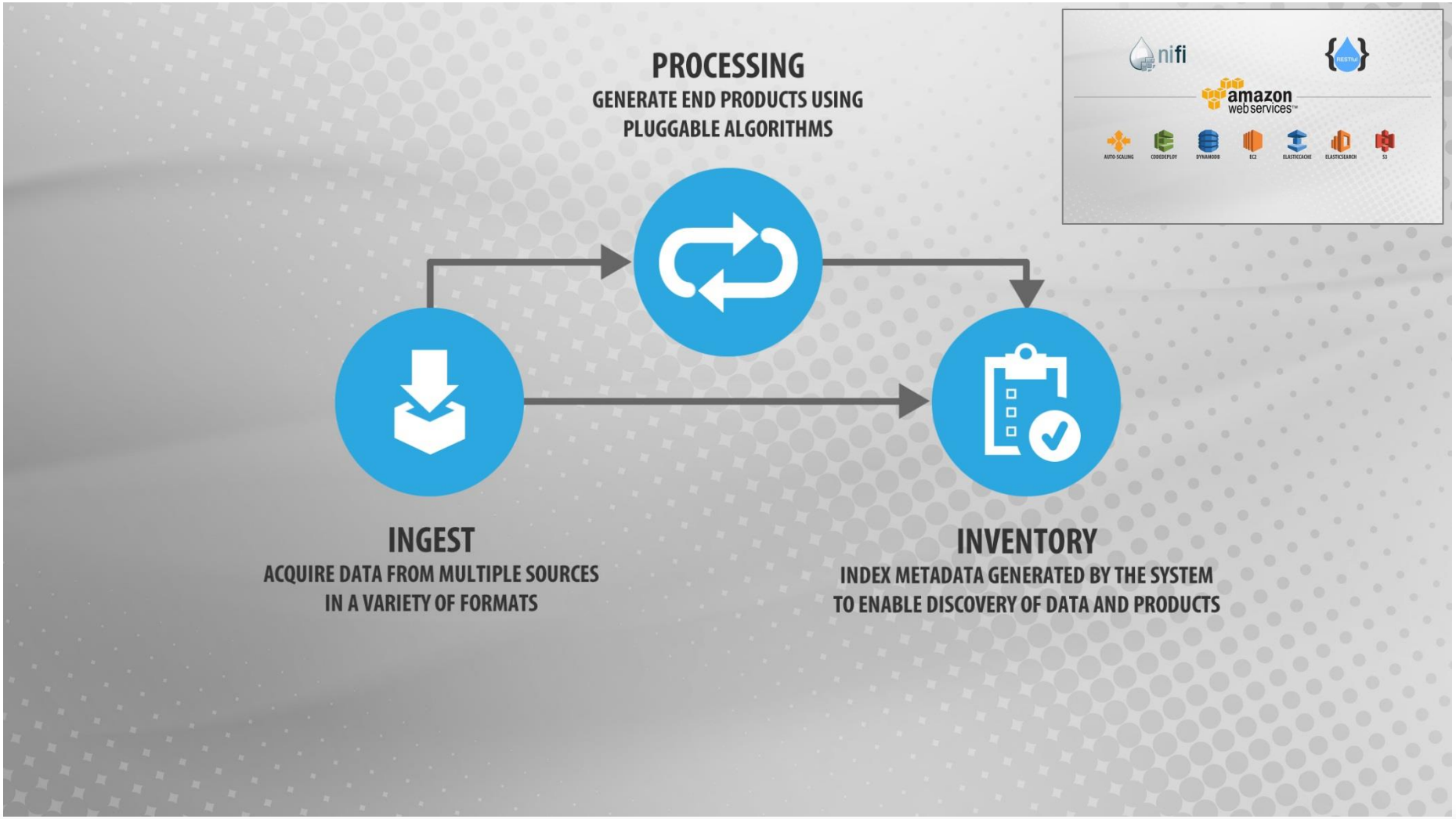
Cloud Based Ingest & Processing Framework (I&PF)

Uses Readily Available Open Source Technologies and Commercial Amazon Cloud Services

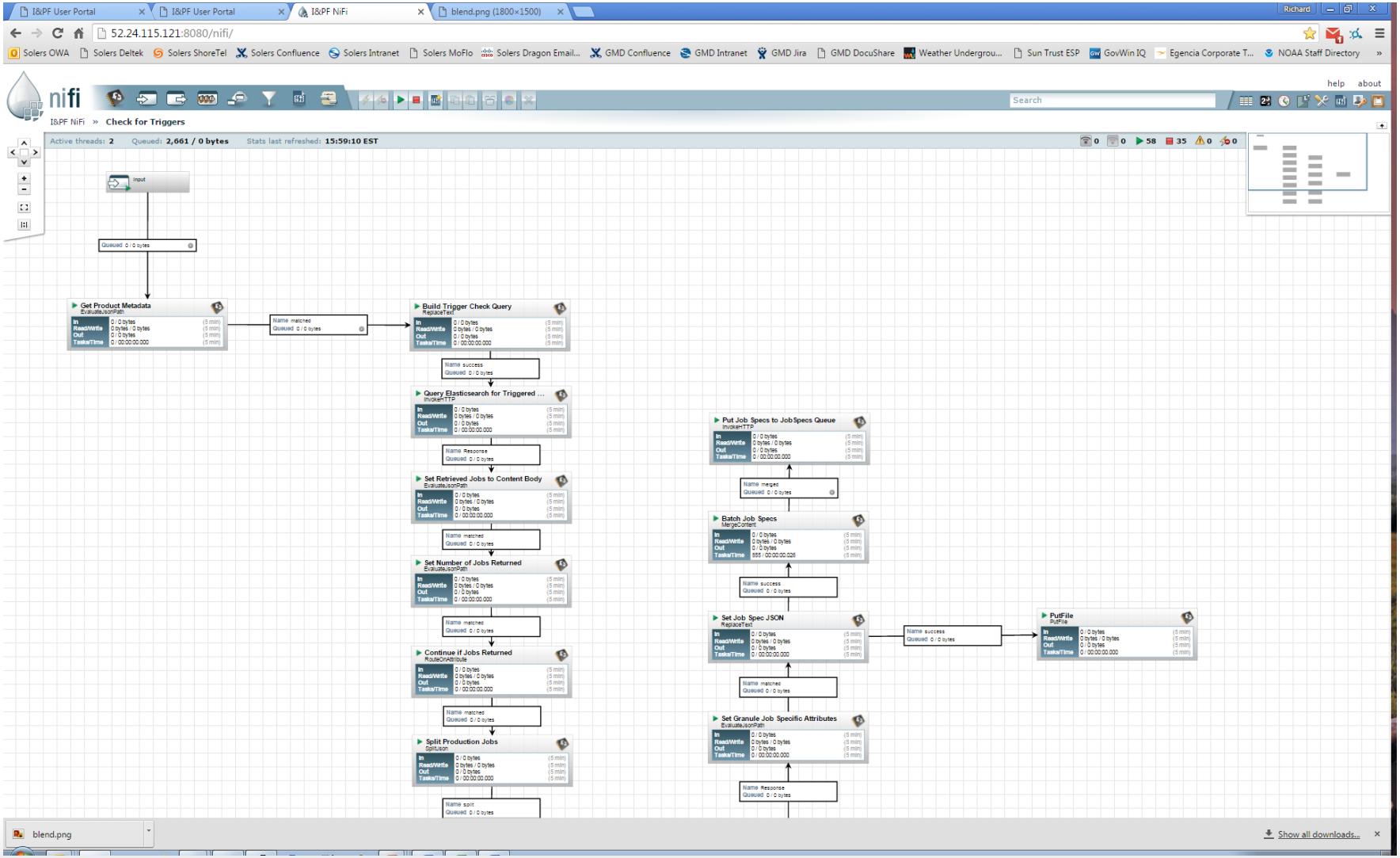
OBJECTIVES:

- ***Enable fast/easy integration*** of data sources, product algorithms, and data consumers within a cloud based workflow (or "data pipeline") framework
- Provide ***easy to use web-based user interfaces*** for discovery and access (for end users), as well as workflow monitoring and management (for algorithm developers and system operators/admins)
- Provide ***RESTful web services*** for other developers, scientists, etc. to ***discover and access*** the ingested/processed ***data and metadata***, for use in other research / engineering initiatives (e.g., developing a new product algorithm)

Cloud Based I&PF
High-Level Architecture



Cloud Based I&PF
Apache NiFi Workflow Engine



Cloud Based I&PF
3 NOAA Proof of Concept Use Cases (Solers IR&D)

➢NOAA S-NPP ATMS and MIRS

- Ingests and inventories Suomi National Polar Partnership (S-NPP) Advanced Technology Microwave Sounder (ATMS) granules
- Generates Microwave Integrated Retrieval System (MIRS) products from the ATMS granules
- Makes ATMS granules and MIRS products searchable and accessible

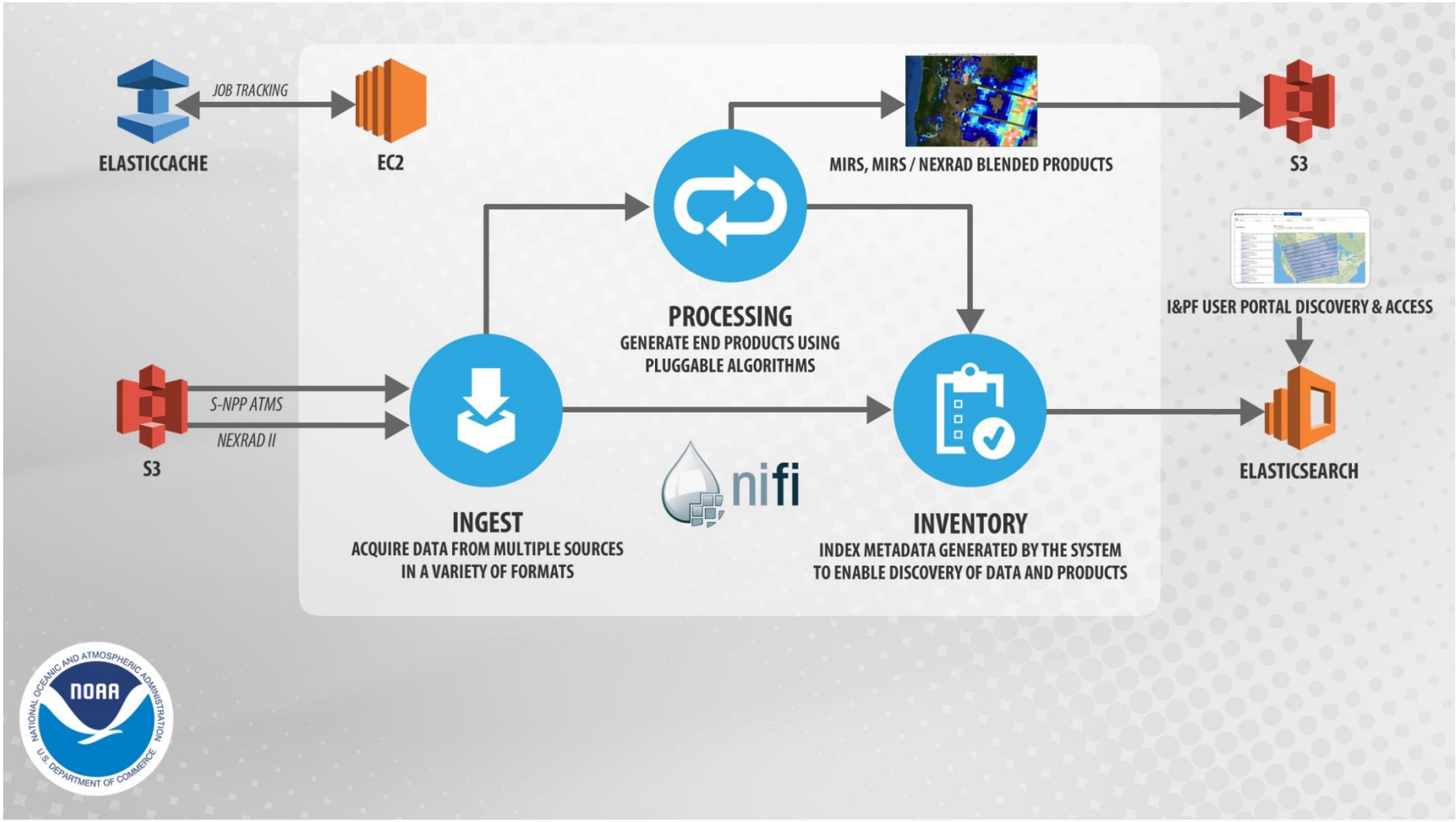
➢NOAA Nexrad II Weather Radar

- Ingests and inventories NOAA Nexrad II Weather Radar data sets that were published on Amazon S3 as part of the NOAA Big Data Project
- Makes NOAA Nexrad II Weather Radar data sets searchable and accessible

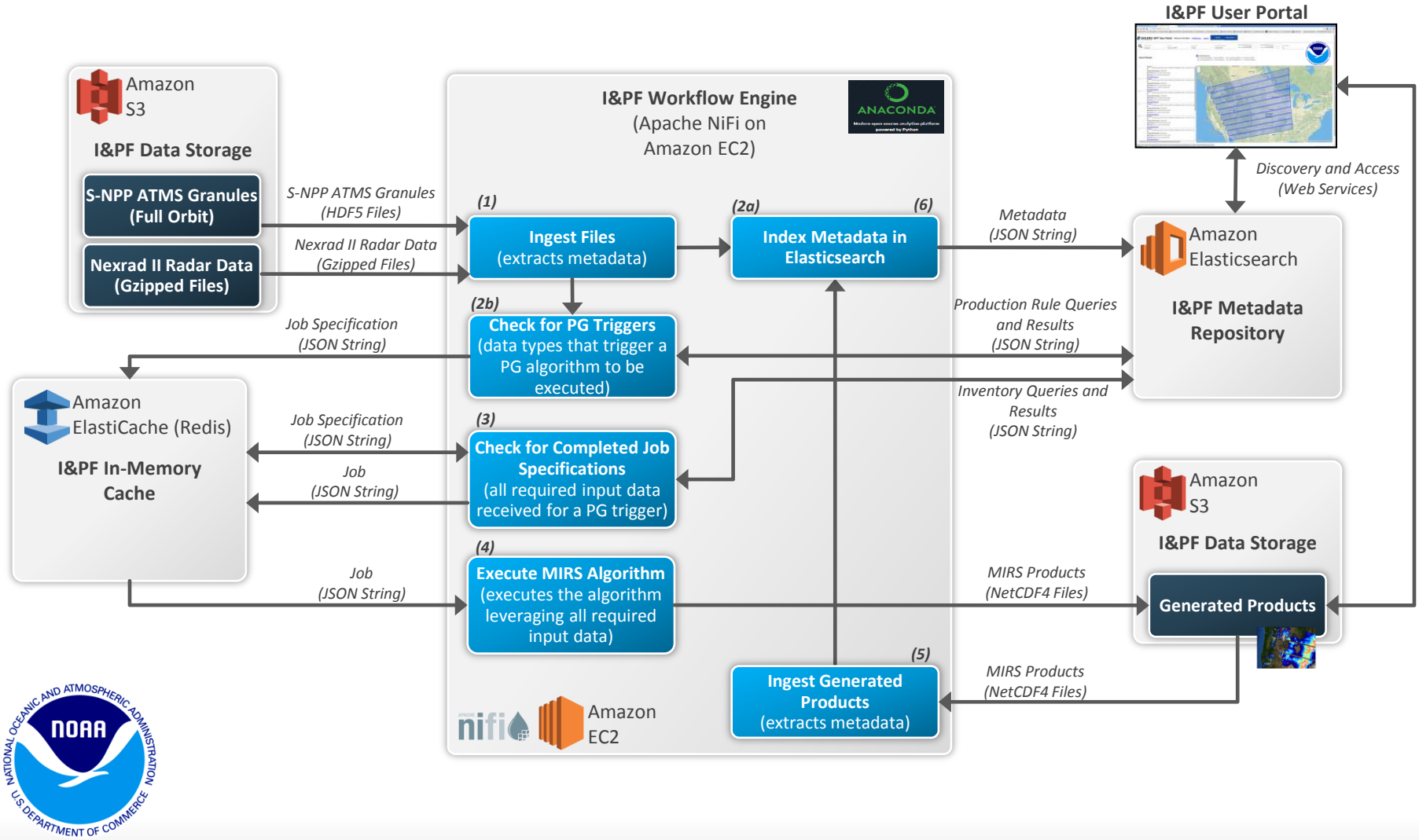
➢MIRS / Nexrad II Blended Product

- Leverages the available MIRS products and NOAA Nexrad II Weather Radar data sets to produce a new blended product that combines the MIRS snow/water data with the Nexrad II radar data over mountainous regions

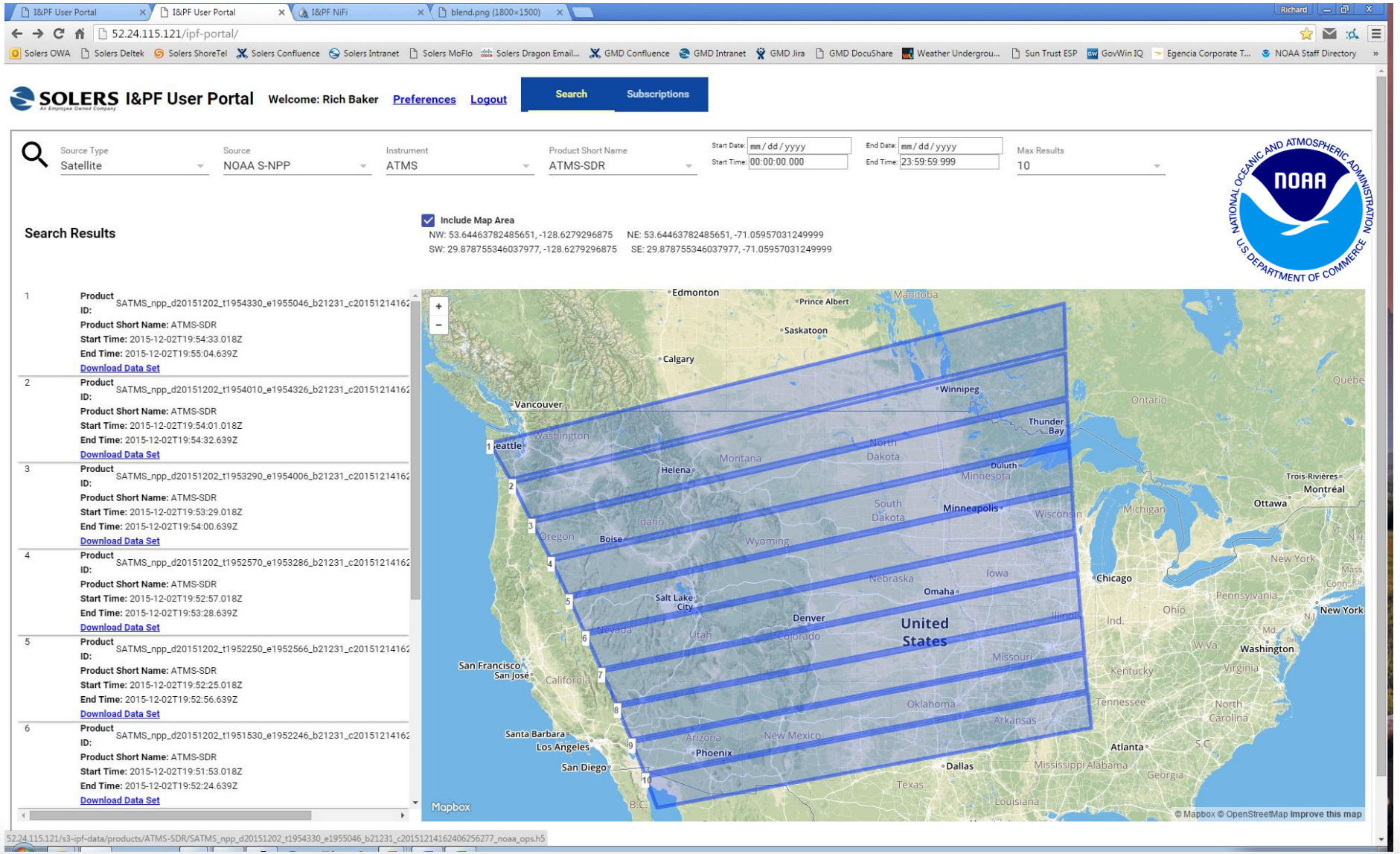
Cloud Based I&PF
NOAA Proof of Concept Architecture



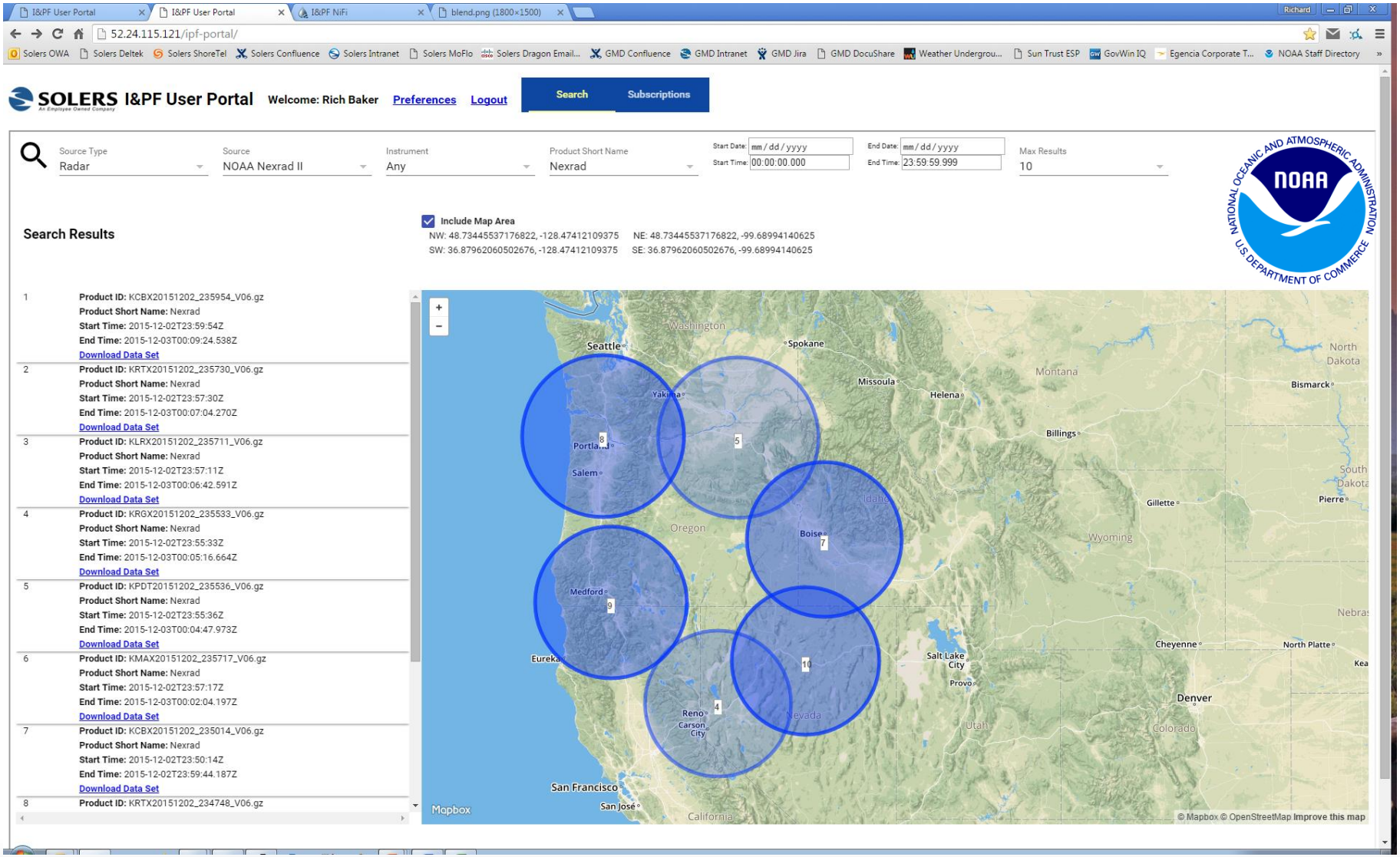
Cloud-Based I&PF NOAA Proof of Concept
NOAA Data Ingest, Processing, and Inventory



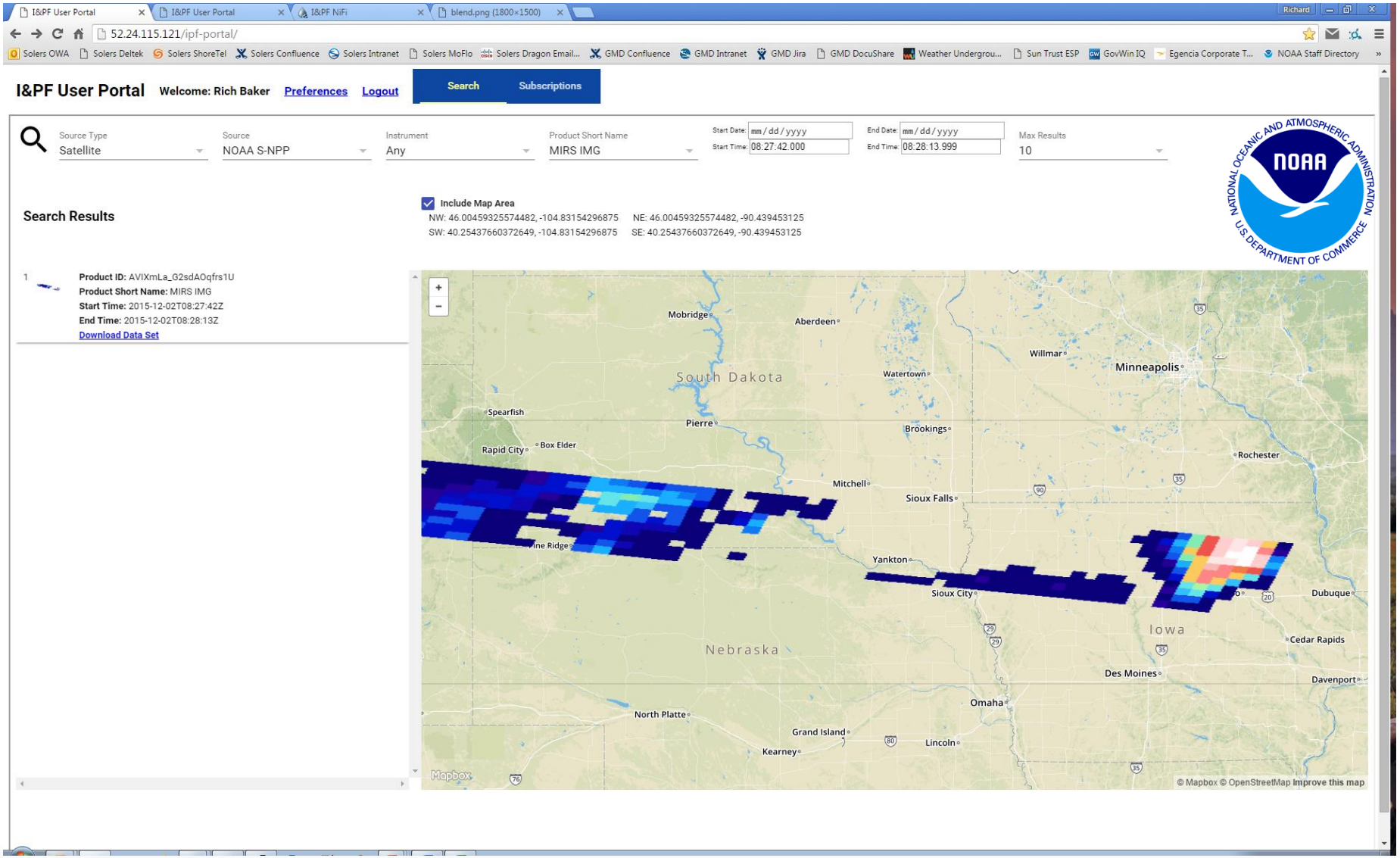
NOAA Data/Products Made Available in the Cloud-Based I&PF: Ingested S-NPP ATMS Granules



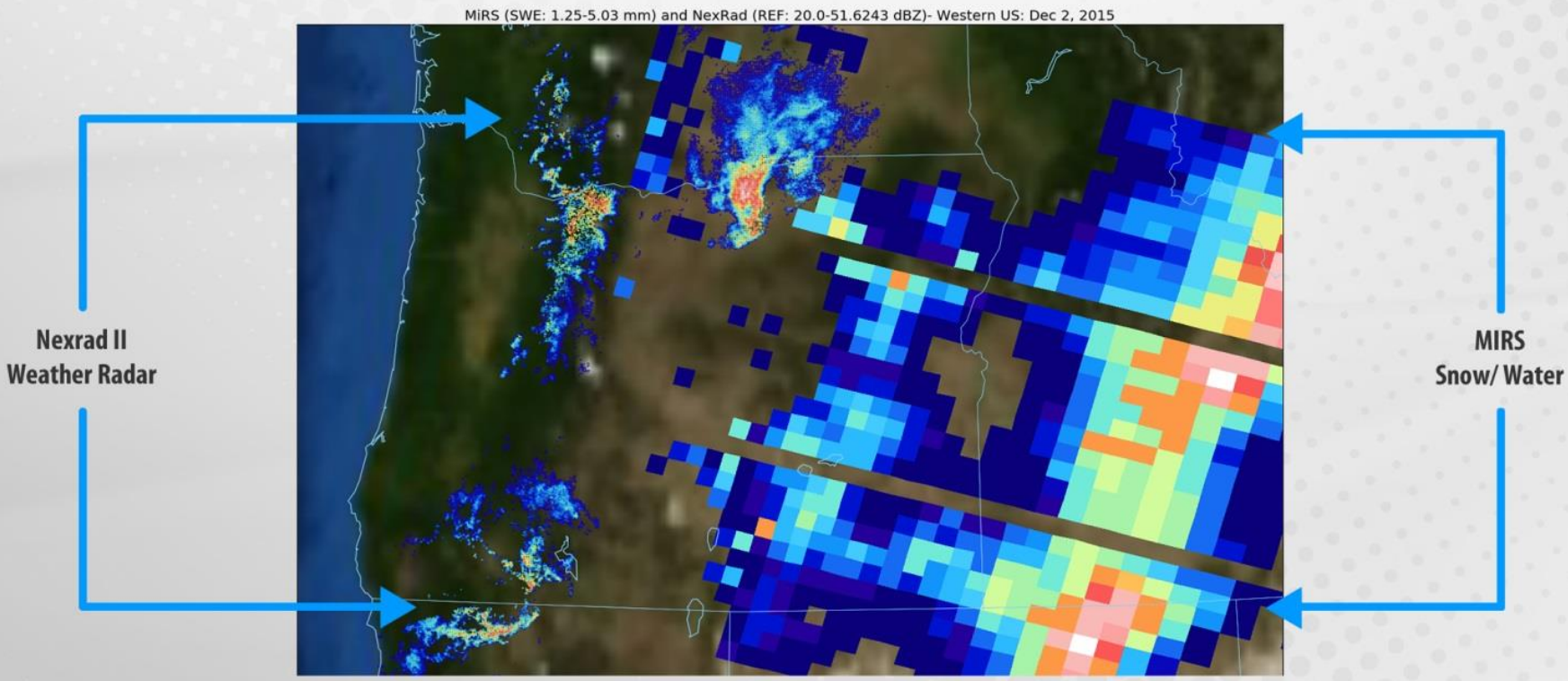
NOAA Data/Products Made Available in the Cloud-Based I&PF: Ingested Nexrad II Weather Radar Data



NOAA Data/Products Made Available in the Cloud-Based I&PF: Generated MIRS Products



NOAA Data/Products Made Available in the Cloud-Based I&PF: MIRS / Nexrad II Blended Product



Cloud Based I&PF
OmniEarth Commercial Project

➢OmniEarth Overview

- OmniEarth utilizes large satellite imagery sets combined with advanced machine learning algorithms to classify land cover for purposes of determining outdoor water budgets at the parcel level
- These budgets aid water agencies in drought-ridden communities in the US to best target water over-users

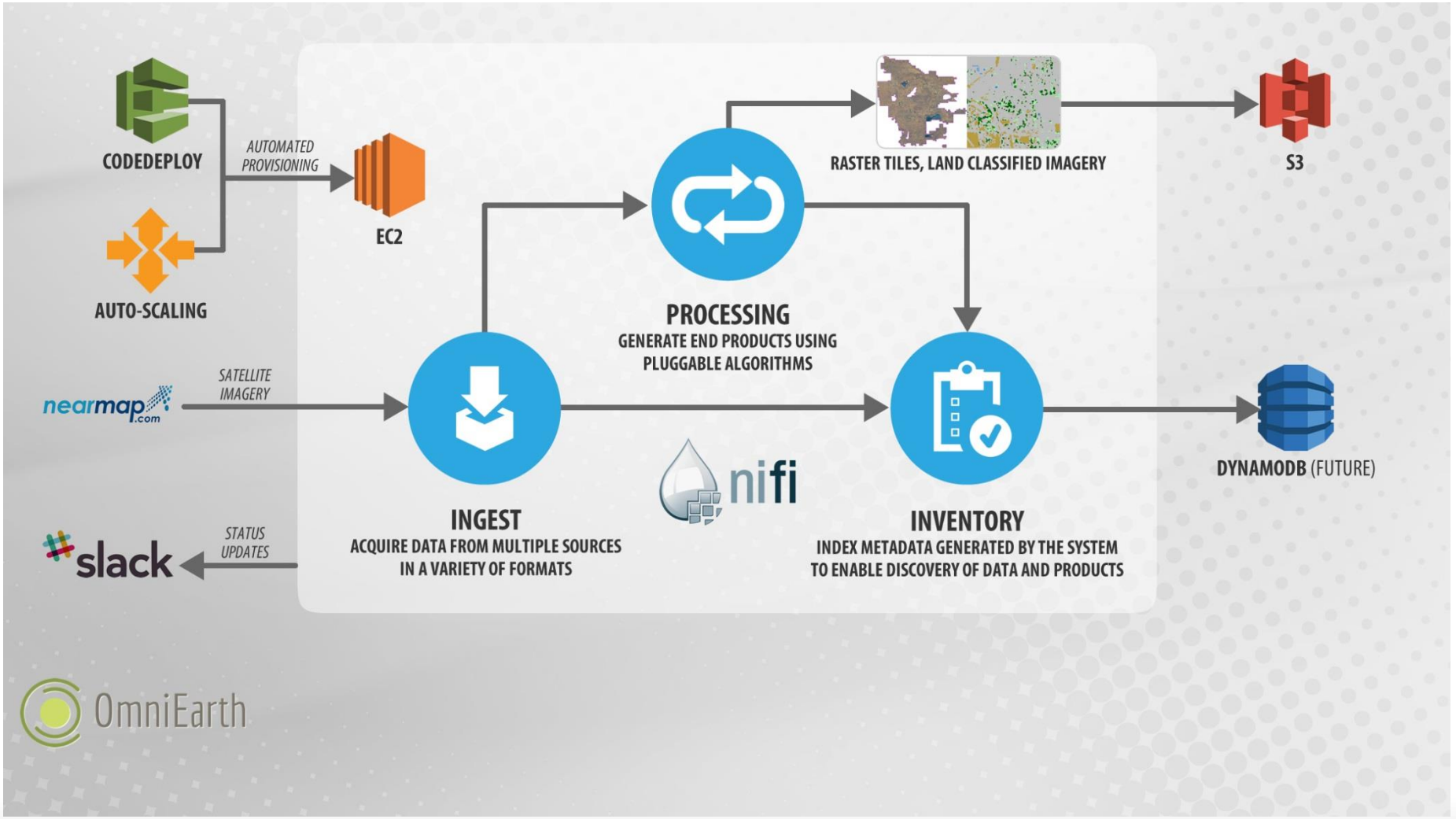
➢Solers' OmniEarth Commercial Project

- Solers has partnered with OmniEarth's Data Scientists to help them utilize the Cloud Based I&PF in order to automate their (previously manual) satellite imagery ingest and land classification algorithm processing activities for their commercial Water Resource Management product

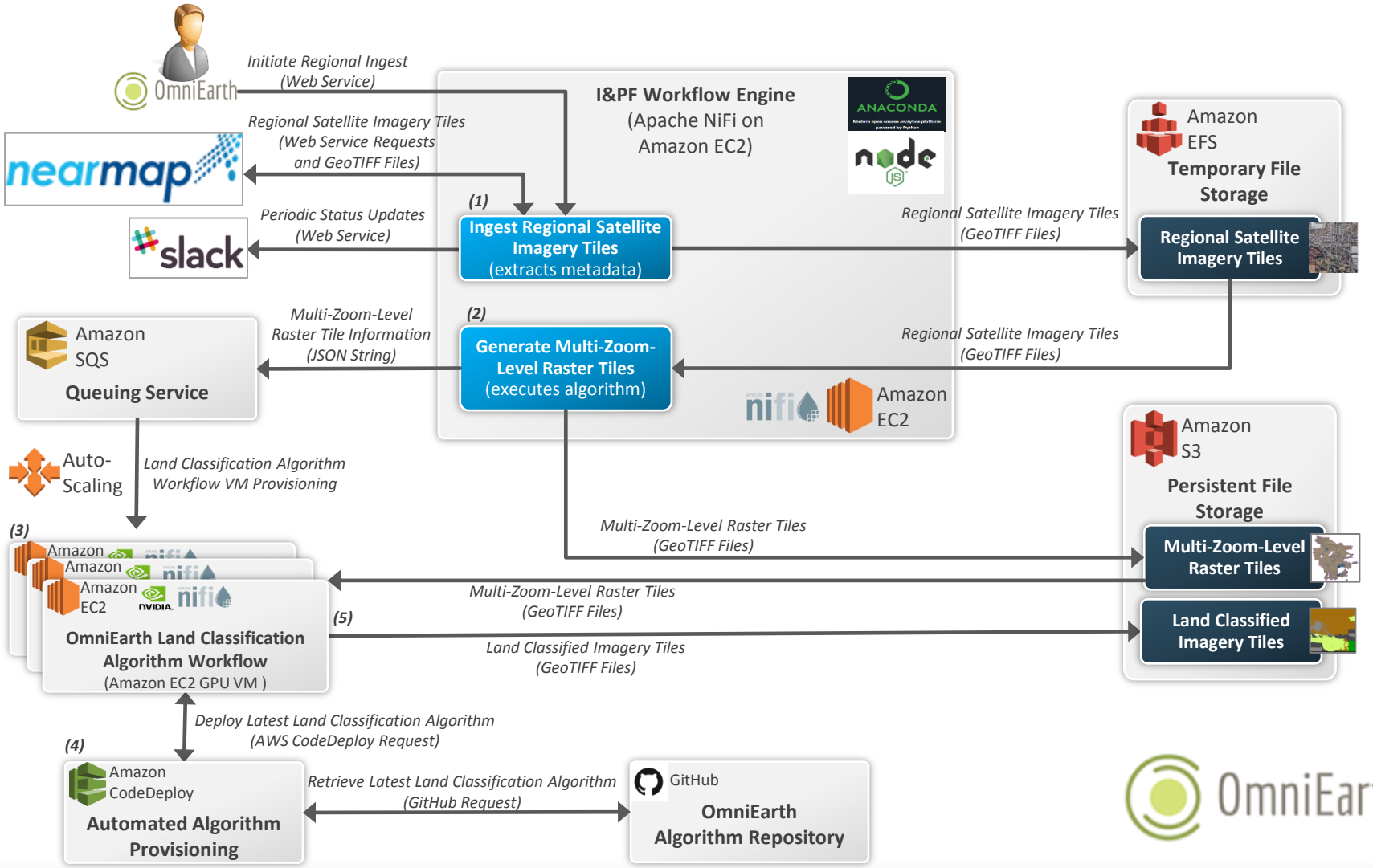
OmniEarth

Water Resource Management Information:
http://water.omniearth.net

Cloud Based I&PF
OmniEarth Commercial Project Architecture

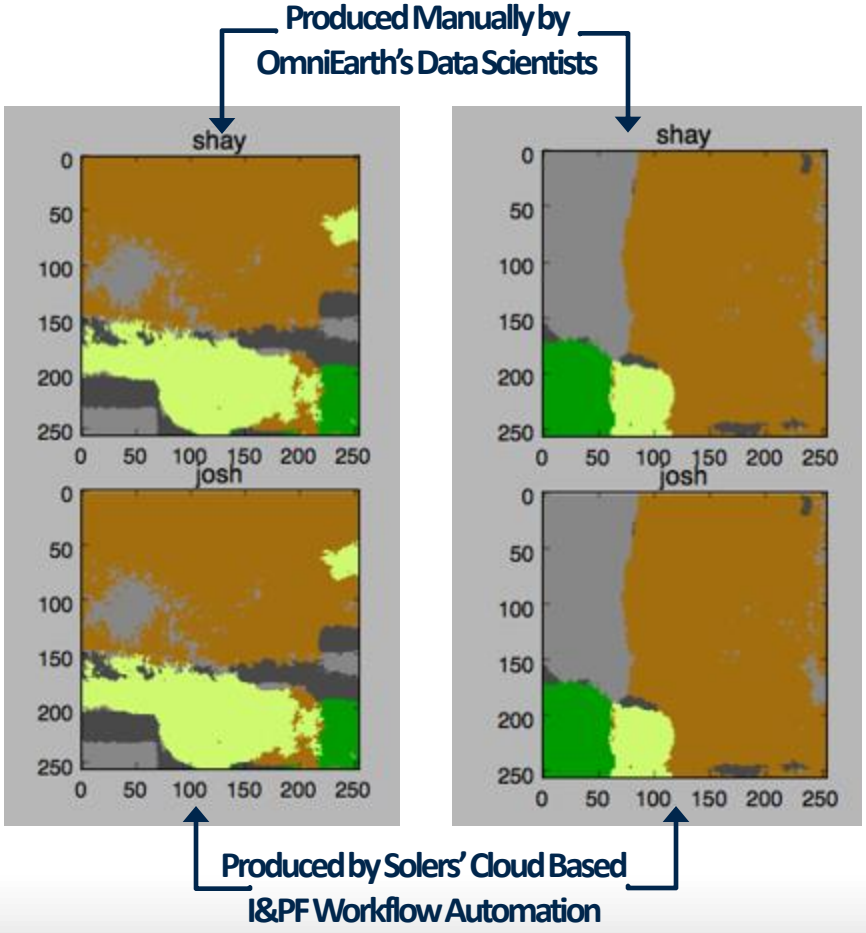


Cloud Based I&PF OmniEarth Commercial Project
Satellite Imagery Ingest and Land Classification Processing



Cloud Based I&PF OmniEarth Commercial Project Automation At Scale, Without Quality Reduction

Solers' Cloud-Based I&PF workflow automation produces land classified imagery tiles for OmniEarth, with the same level of precision and accuracy as those produced manually



OmniEarth

SOLERS
An Employee Owned Company

Cloud Based I&PF OmniEarth Commercial Project Outcomes and Benefits for OmniEarth

➢Automation and Efficiency

- Automates their previously manual satellite data ingest and land classification processing activities
- Reduces the time to perform these activities by an order of magnitude (days to hours)
- Allows OmniEarth's Data Scientists to focus on improving their algorithms and training models, instead of manually running and watching over the satellite data ingest and land classification algorithm execution

➢Tailored to the Data Scientist Needs

- Simple web service interface to initiate the workflow, based upon customer needs (e.g., specific regions of interest)
- Periodic monitoring/alerting of workflow status using a tool that is already heavily used and familiar to them (Slack)

OmniEarth

Cloud Based I&PF
Future Utilities and Benefits

- **Development, integration, and test environment for Government (e.g., NOAA, NASA) satellite ground systems**
  - Perform R&D and Cal/Val of new product algorithms for multiple satellites/platforms
  - Scalable cloud-based framework that avoids on-premise infrastructure costs (pay just for the services that you need/use)
  - Automation at scale with interfaces tailored to science algorithm developer and data scientist needs, helping to reduce the Research to Operations (R2O) timeline
- **Ingest and processing framework for commercial small satellite startup companies**
  - Enable them to quickly get their satellite data ingested, processed, and available to users via a scalable cloud-based workflow or “data pipeline” framework, without requiring on-premise infrastructure

AMS 2017 Theme: Observations Lead The Way

- **Your view on the greatest observational needs for your discipline in general**
  - Promotion of cloud-based platforms/frameworks for NOAA (FedRAMP approved, such as AWS GovCloud) to perform value-added capabilities with available observation data sets that are being published to cloud storage services (such as S3) as part of ongoing cloud-related initiatives (such as the NOAA Big Data Project)
  - Index/catalog the data for discovery and access, and leverage cloud services to perform other value-added capabilities such as product generation, data assimilation, re-processing, etc. (more than just storing the data)

Questions